CSE472 Assignment 2 Report

Submitted by:
ID: 1805012

Instructions

1. To run a particular dataset, we need to import the respective preprocessor. The preprocessors are named as “preprocessor_*.py”. Each one of them have a function named “preprocess()”.
2. Using the preprocess() function, we will get X_train, y_train, X_test, y_test datasets.
3. These datasets can be directly used along with LogisticRegression or AdaBoost. Changing the preprocessor will change the dataset.

```
from preprocess_adult import preprocess

X_train_top_k, y_train, X_test_top_k, y_test = preprocess()
```

Experimental Results

Logistic Regression

Telco Customer Churn

Performance measure	Training	Test
Accuracy	**0.772**	**0.735**
True positive rate (sensitivity, recall, hit rate)	**0.83**	**0.81**
True negative rate (specificity)	**0.71**	**0.66**
Positive predictive value (precision)	**0.74**	**0.70**
False discovery rate	**0.26**	**0.30**

F1 score	**0.79**	**0.75**

Adult - UCI Machine Learning Repository

Performance measure	**Training**	**Test**
Accuracy	**0.7436**	**0.7471**
True positive rate (sensitivity, recall, hit rate)	**0.55**	**0.55**
True negative rate (specificity)	**0.94**	**0.94**
Positive predictive value (precision)	**0.90**	**0.91**
False discovery rate	**0.10**	**0.09**
F1 score	**0.68**	**0.68**

Credit Card Fraud Detection

Performance measure	**Training**	**Test**
Accuracy	**0.9261**	**0.92525**
True positive rate (sensitivity, recall, hit rate)	**0.85**	**0.86**
True negative rate (specificity)	**1.00**	**1.00**

Positive predictive value (precision)	**1.00**	**1.00**
False discovery rate	**0.00**	**0.00**
F1 score	**0.92**	**0.92**

AdaBoost

Telco Customer Churn

Number of Boosting Rounds	Training	Test
5	**0.7451**	**0.7246**
10	**0.7718**	**0.7377**
15	**0.7435**	**0.7087**
20	**0.7537**	**0.7329**

Adult - UCI Machine Learning Repository

Number of Boosting Rounds	Training	Test
5	**0.7887**	**0.7918**
10	**0.49**	**0.50**

15	**0.48**	**0.49**
20	**0.48**	**0.49**

Credit Card Fraud Detection

Number of Boosting Rounds	**Training**	**Test**
5	**0.7508**	**0.7690**
10	**0.5573**	**0.5717**
15	**0.6592**	**0.6757**
20	**0.6448**	**0.6580**